Random walks on quasirandom graphs

Ben Barber *   Eoin Long †

November 15, 2012

Abstract

Let G be an n-vertex quasirandom graph with $\rho\binom{n}{2}$ edges, and let W be a random walk on G of length βn^2. Let G' be the graph obtained from G by deleting the edges traversed by W. We show that (for fixed ρ and β) with high probability G' is quasirandom with $(e^{-2\beta/\rho}+o(1))\rho\binom{n}{2}$ edges. We also obtain a similar result when the random walk is replaced by a random homomorphism of a fixed tree with maximum degree $c\sqrt{\log n}$ for a small constant c. This answers a question of Böttcher, Hladký, Piguet and Taraz that arose in the context of tree packing.

1 Introduction

Given a graph G and sets $A, B \subseteq V(G)$ let $E_G(A,B) = \{(a,b) \in A \times B : ab \in E(G)\}$ and let $e_G(A,B) = |E_G(A,B)|$. A graph G with n vertices and $\rho\binom{n}{2}$ edges is ϵ-*quasirandom* if

$$|e_G(A,B) - \rho|A||B|| < \epsilon|A||B|$$

for all sets $A, B \subseteq V(G)$ with $|A|, |B| \geq \epsilon n$. Thus a quasirandom graph resembles a random graph with the same density, provided we do not look too closely. Quasirandom graphs were introduced by Thomason [10] and have come to play a central role in probabilistic and extremal graph theory. The reader is referred to the excellent survey article by Krivelevich and Sudakov [5] for further details.

Suppose that G' is a random subgraph of G in which each edge is included independently with a fixed probability p. It is easy to see that, for any $\eta > \epsilon$, the graph G' is with high probability η-quasirandom, provided n is large enough. Our main result is that a similar conclusion holds when the subgraph G' of G is chosen according to another natural distribution.

A walk W on G of length l consists of a sequence of vertices $W = W_0 \dots W_l$ where W_iW_{i+1} is an edge of G for all $i < l$. A *random walk* W of length l on G is obtained by choosing a start vertex W_0 of G from some initial distribution (typically this is either specified precisely or chosen uniformly at random) and then, for each $i < l$, choosing W_{i+1} uniformly at random from the neighbours of W_i, with each choice made independently.

*Department of Pure Mathematics and Mathematical Statistics, Centre for Mathematical Sciences, Wilberforce Road, Cambridge, CB3 0WB, UK. b.a.barber@dpmms.cam.ac.uk

†School of Mathematical Sciences, Queen Mary University of London, UK. e.p.long@qmul.ac.uk

Böttcher, Hladký, Piguet and Taraz [3] recently asked the following question. Let W be a random walk of length βn^2 on G and let G' be the graph obtained from G by deleting the edges of W. Is G' quasirandom with high probability? This question arose in the context of a random tree embedding procedure which sought to find disjoint copies of a given collection of trees in the complete graph K_n. Böttcher, Hladký, Piguet and Taraz noticed that if removing randomly embedded copies of a small number of such trees from a quasirandom graph G preserves quasirandomness, it may be possible to nibble (or iteratively remove) many disjoint trees from K_n.

Our main result is that G' *is* quasirandom with high probability. We prove two versions of this statement. The first applies when the minimum degree of G is reasonably large.

Theorem 1 (Bounded minimum degree). *Let $\beta, \epsilon, \rho, \eta > 0$ with $\eta > \epsilon$ and let $\gamma = C\epsilon^{1/4}$ for some absolute constant $C > 0$. Let G be an n-vertex ϵ-quasirandom graph with $\rho\binom{n}{2}$ edges and minimum degree at least γn, and let W be a random walk on G of length βn^2. Then, with probability $1-o(1)$, the graph G' is η-quasirandom with $(e^{-2\beta/\rho} + o(1))\rho\binom{n}{2}$ edges.*

In fact, our proof gives more than this. It shows that with high probability $e_{G'}(A, B) = (e^{-2\beta/\rho} + o(1))e_G(A, B)$ for all sets $A, B \subseteq V(G)$ with $e_G(A, B) \gg n$. Thus the subgraph spanned by W looks like a random subgraph of G in a very strong sense.

The bound on the minimum degree of G means that G is well-connected and allows us to take advantage of a general result on the rate of convergence of a random walk. For general quasirandom graphs such results do not hold, and at the start of Section 3 we give an example of a poorly-connected graph for which the number of edges in G' can take very different values with positive probability, depending on ϵ but not on n. To obtain a version of Theorem 1 for such graphs we must therefore allow our probability to depend on ϵ as well as n. We write $o_\epsilon(1)$ for a quantity that is less than $f(\epsilon)$ for n sufficiently large, where $f(\epsilon) \to 0$ as $\epsilon \to 0$.

Theorem 2 (General case). *Given $\beta, \rho, \eta > 0$ there exists $\epsilon > 0$ such that the following holds. Let G be an n-vertex ϵ-quasirandom graph with $\rho\binom{n}{2}$ edges and let W be a random walk on G of length βn^2 starting at any vertex W_0 of G with degree in $[(\rho-\epsilon)n, (\rho+\epsilon)n]$. Then, with probability $1 - o_\epsilon(1)$, the graph G' is η-quasirandom with $(e^{-2\beta/\rho} + o_\epsilon(1))\rho\binom{n}{2}$ edges.*

It is easily seen that there are at least $(1-2\epsilon)n$ choices for such a vertex W_0 in G (see Proposition 5), so a vertex of G selected uniformly at random satisfies the conditions of Theorem 2 with high probability.

The proof of Theorem 1 is given in Section 2. We then extend the proof to the general case of Theorem 2 in Section 3 with some additional ideas. In Section 4 we discuss the extension of our methods to random homomorphisms of general trees.

Since we will only prove asymptotic results we make a number of simplifying assumptions. We assume ϵ is sufficiently small compared to the other parameters, and are only interested in statements for n sufficiently large. We omit notation indicating the taking of integer parts, and ignore questions of divisibility when breaking walks into pieces of a given size.

2 Bounded minimum degree

In our proof of Theorem 1 we will take the following alternative perspective on the construction of a random walk W. For each vertex v, let L_v be an infinite list of neighbours of v with each entry selected independently and uniformly at random. As before, choose W_0 from some given distribution. Then, at every stage i, if the walk has just made its jth visit to vertex v, let W_{i+1} be the jth element of L_v. It is easy to see that this gives the same distribution on random walks W as described in Section 1.

We will prove Theorem 1 in two stages. In Section 2.1, we show that with high probability W visits each vertex of G about as often as we expect. Observe that, using our alternative description of W, this tells us (roughly) how many elements from each list L_v we used in the construction of W. In Section 2.2 we show that after removing the edges of G corresponding to these elements from the top of the lists L_v we are left with a graph which strongly resembles a random subgraph of G of the appropriate density.

2.1 The number of visits to each vertex

To begin this subsection we recall some useful facts. A random walk W on a graph G is a Markov chain with transition matrix P given by

$$P_{uv} = \begin{cases} 1/d(u) & \text{if } uv \in E(G); \\ 0 & \text{if } uv \notin E(G). \end{cases}$$

Thus P is a normalised version of the adjacency matrix A where each row has been scaled by the degree of the corresponding vertex. The eigenvalues of P are all real; let these be $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ and write $\lambda = \max(|\lambda_2|, |\lambda_n|)$. The first eigenvalue λ_1 of P is always equal to 1 and has a corresponding eigenvector $\pi = (\pi_v)$ given by $\pi_v = \frac{d(v)}{2e(G)}$. This vector π is called the *stationary distribution* of the walk W. It is well-known (for example see [6]) that if G is connected and non-bipartite then, for any initial distribution of W_0, the distribution of W_i converges to π as $i \to \infty$ (i.e. $\mathbb{P}(W_i = v) \to \pi_v$ as $i \to \infty$ for each v). The following standard result, which can read out of Jerrum and Sinclair [4], gives control on the *rate* of this convergence.

Lemma 3. *For any n-vertex graph G with minimum degree at least γn and any initial distribution on W_0, we have*

$$\max_{v \in V(G)} |\mathbb{P}(W_i = v) - \pi_v| \leq c_\gamma \lambda^i,$$

for some c_γ depending on γ.

Now if G is a regular ϵ-quasirandom graph then λ is small on the scale of ϵ. (This is because the 'spectral gap' of a quasirandom graph is large [2], and P is a scalar multiple of A when G is regular.) For a general ϵ-quasirandom graph this need not be true: for example, if G contains a small connected component, then $\lambda = 1$ (the 1-eigenspace is spanned by the stationary distribution of each connected component of G). Similarly, λ can be very close to 1 if there is a small set of vertices that is only weakly connected to the rest of the graph. However, a lower bound on the minimum degree of G is enough to recover an upper bound on λ.

Lemma 4. *Let G be an n-vertex ϵ-quasirandom graph with minimum degree at least γn where $\rho, \gamma \geq C\epsilon^{1/4}$ for some absolute constant $C > 0$. Then, for n sufficiently large, $\lambda \leq 1/2$.*

Before proving Lemma 4, we make the following simple observation about quasirandom graphs which we will use repeatedly.

Proposition 5. *Let G be an n-vertex ϵ-quasirandom graph with $\rho\binom{n}{2}$ edges, and let X be a set of vertices with $|X| \geq \epsilon n$. Let $Y = \{v \in V(G) : |e(v, X) - \rho|X|| \geq \epsilon|X|\}$. Then $|Y| < 2\epsilon n$.*

Proof. We have $Y = Y^- \cup Y^+$ where

$$Y^+ = \{v \in V(G) : e(v, X) \geq \rho|X| + \epsilon|X|\},$$
$$Y^- = \{v \in V(G) : e(v, X) \leq \rho|X| - \epsilon|X|\}.$$

Clearly

$$\left|e_G(X, Y^+) - \rho\,|X|\,\left|Y^+\right|\right| \geq \epsilon\,|X|\,\left|Y^+\right|,$$
$$\left|e_G(X, Y^-) - \rho\,|X|\,\left|Y^-\right|\right| \geq \epsilon\,|X|\,\left|Y^-\right|.$$

But then, since G is ϵ-quasirandom and $|X| \geq \epsilon n$, we must have $|Y^+|, |Y^-| < \epsilon n$. ☐

In particular, taking $X = V(G)$ there are at least $(1 - 2\epsilon)n$ vertices v of G with $|d(v) - \rho n| \leq \epsilon n$. We will call such vertices *balanced*.

Proof of Lemma 4. The proof follows a well-known argument (see for example [2]). We first estimate the number of labelled copies of C_4 in G, and then evaluate the trace of P^4 in two different ways. Note that the implicit constants in our use of $O(\cdot)$ notation here are absolute.

The number of labelled copies of C_4 in G is

$$\begin{aligned}
C_4(G) &= 2 \sum_{u \in V(G)} \sum_{v \in V(G)} \binom{|N(u) \cap N(v)|}{2} \\
&= 2 \cdot (1 + O(\epsilon))n \cdot (1 + O(\epsilon))n \cdot \binom{(\rho + O(\epsilon))^2 n}{2} + O(\epsilon)n^2\binom{n}{2} \\
&= (\rho + O(\epsilon))^4 n^4 + O(\epsilon)n^4 \\
&= \big(1 + O\big(\epsilon/\rho^4\big)\big)\, \rho^4 n^4.
\end{aligned}$$

where the main term here accounts for balanced vertices u and v with close to $\rho^2 n$ common neighbours, and the error term bounds the contribution to the sum from each other pair by $\binom{n}{2}$.

Now the trace of P^4 is a weighted sum of the closed walks of length 4 in G, where the weight of the closed walk $uvwx$ is $1/(d(u)d(v)d(w)d(x))$. Thus

$$\begin{aligned}
\sum_{v \in V(G)} (P^4)_{vv} &= \frac{(1 + O(\epsilon/\rho^4))\rho^4 n^4}{((\rho + O(\epsilon))n)^4} + \frac{O(\epsilon)n^4}{(\gamma n)^4} + \frac{O\big(n^3\big)}{(\gamma n)^4} \\
&= 1 + O\big(\epsilon/\rho^4\big) + O\big(\epsilon/\gamma^4\big) + O\big(1/(\gamma^4 n)\big),
\end{aligned}$$

where the main term counts the contribution from 4-cycles containing only balanced vertices and the error terms account for the contributions from 4-cycles with at least one unbalanced vertex and from closed walks of length 4 which are not 4-cycles respectively. (The lower bound on the minimum degree of G gives an upper bound of $1/(\gamma n)^4$ for the weight of any one walk.) But we also have

$$\sum_{v\in V(G)} (P^4)_{vv} = \sum_{i=1}^{n} \lambda_i^4 = 1 + \sum_{i=2}^{n} \lambda_i^4,$$

from which it follows that

$$\lambda^4 \leq \sum_{i=2}^{n} \lambda_i^4 = O\big(\epsilon/\rho^4\big) + O\big(\epsilon/\gamma^4\big) + O\big(1/(\gamma^4 n)\big) \leq 1/16,$$

for $\rho, \gamma \geq C\epsilon^{1/4}$ and n sufficiently large. □

For the next lemma we will need to approximate one probability measure by another on the same space. Given a finite probability space Ω, the *total variation distance* between two probability measures μ_1 and μ_2 is defined by

$$d_{TV}(\mu_1, \mu_2) = \frac{1}{2} \sum_{\omega\in\Omega} |\mu_1(\omega) - \mu_2(\omega)|.$$

This is the amount of probability mass that would have to be moved to turn one distribution into the other.

Combining Lemma 3 with Lemma 4, it is easy to see that the total variation distance between W_t and a vertex sampled from the stationary distribution is small when t is moderately large. In fact, we get much more.

Let $L = (\log n)^2$ and let $K = \beta n^2/L$. Given $i < L$, let $W^{(i)}$ denote the subsequence of W obtained by starting from W_i and taking L steps at a time: that is, $W^{(i)} = (W_1^{(i)} \dots, W_K^{(i)})$ where $W_j^{(i)} = W_{i+jL}$ for all $j < K$. For each $v \in V(G)$ let $X_v^{(i)}$ denote the random variable which counts the number of times $W^{(i)}$ visits v. Our next lemma shows that with high probability $X_v^{(i)}$ is close to its mean.

Lemma 6. *Let G be a graph satisfying the conditions of Lemma 4 and $v \in V(G)$. Then we have*

$$\mathbb{P}\left(\left|X_v^{(i)} - K\pi_v\right| \geq \sqrt{\frac{8\log n}{K\pi_v}} K\pi_v\right) = O\big(n^{-3}\big).$$

Proof. Let $\mu = \pi^K$ be the K-fold product measure of π on $V(G)^K$; that is, $\mu(w) = \prod_{i=1}^K \pi_{w_i}$ for $w \in V(G)^K$. By Lemma 3 and Lemma 4, we have

$$\begin{aligned}
\mathbb{P}\left(W^{(i)} = w\right) &= \mathbb{P}\left(W_1^{(i)} = w_1\right)\mathbb{P}\left(W_2^{(i)} = w_2 | W_1^{(i)} = w_1\right)\cdots\mathbb{P}\left(W_K^{(i)} = w_K | W_{K-1}^{(i)} = w_{K-1}\right)\\
&= \left(\pi_{w_1} + O\Big(2^{-(\log n)^2}\Big)\right)\left(\pi_{w_2} + O\Big(2^{-(\log n)^2}\Big)\right)\cdots\left(\pi_{w_K} + O\Big(2^{-(\log n)^2}\Big)\right)\\
&= \left(\pi_{w_1} + O\big(n^{-6}\big)\right)\left(\pi_{w_2} + O\big(n^{-6}\big)\right)\cdots\left(\pi_{w_K} + O\big(n^{-6}\big)\right)\\
&= \left(1 + O\big(n^{-3}\big)\right)\mu(w),
\end{aligned}$$

since $\frac{\gamma}{\rho n} \le \pi_v \le \frac{1}{\rho n}$ for all v and $K = O(n^2)$. Summing over all w then gives that

$$d_{TV}(\mathbb{P}, \mu) = O(n^{-3}),$$

where $\mathbb{P}$ is the measure on $V(G)^K$ induced by $W^{(i)}$. Now let

$$A = \left\{ w \in V(G)^K : |X_v^{(i)}(w) - K\pi_v| \ge \sqrt{2 \log n K \pi_v} \right\}.$$

By Chernoff's inequality (see [1, A.1.11 and A.1.13]),

$$\mu(A) \le 2e^{-(4+o(1))\log n} = O(n^{-3}).$$

Since $\mathbb{P}(A) \le \mu(A) + d_{TV}(\mathbb{P}, \mu)$, the result follows. □

Now let $X_v = \sum_{i=0}^{L-1} X_v^{(i)}$ be the number of visits W makes to vertex v. Observing that $LK\pi_v = \beta n^2 \pi_v = (1 + \frac{1}{n-1})\frac{\beta}{\rho} d(v)$, we obtain the following corollary by summing over i and v.

Corollary 7. *Let $\beta, \epsilon, \rho, \gamma > 0$ with $\rho, \gamma \ge C\epsilon^{1/4}$ for some absolute constant $C > 0$. Let G be an n-vertex ϵ-quasirandom graph with minimum degree at least γn, and let W be a random walk on G of length βn^2. Then*

$$\mathbb{P}\left(\left| X_v - \frac{\beta}{\rho} d(v) \right| \ge \sqrt{\frac{8 \log n}{K \pi_v}} \frac{\beta}{\rho} d(v) \text{ for some } v \right) = O(n^{-1}). \quad \square$$

Thus with high probability the number of visits W makes to each $v \in V(G)$ is $\left(\frac{\beta}{\rho} + o(1)\right) d(v)$.

2.2 The number of edges removed from subgraphs

Recall that by the definition of the L_v, for each $v \in V(G)$ and $u \in L_v$ we have $uv \in E(G)$. Thus each entry of a list L_v corresponds to an edge of G. For every $\nu \ge 0$, we let $G_{list}(\nu)$ denote the random subgraph of G obtained by deleting the edges of G corresponding to the first $\nu d(v)$ entries in each list L_v. By Corollary 7, we have that $G_{list}(\frac{\beta}{\rho} + o(1)) \subseteq G' \subseteq G_{list}(\frac{\beta}{\rho} - o(1))$ with high probability. It therefore suffices to study the edge distribution of $G_{list}(\nu)$.

We will show that, on large scales, $G_{list}(\nu)$ has the same edge distribution properties as a random graph of the appropriate density. Our next lemma calculates this density.

Lemma 8. *The probability that an edge of G is retained in $G_{list}(\nu)$ is $e^{-2\nu} + o(1)$.*

Proof. For the edge uv to be retained in $G_{list}(\nu)$, v must not appear in the first $\nu d(u)$ entries of L_u, and u must not appear in the first $\nu d(v)$ entries of L_v. Hence the probability that uv is retained in $G_{list}(\nu)$ is

$$\left(1 - \frac{1}{d(u)}\right)^{\nu d(u)} \left(1 - \frac{1}{d(v)}\right)^{\nu d(v)} = e^{-2\nu} + o(1),$$

since $d(u)$, $d(v) \ge \gamma n$. □

To show that the number of edges retained in any subgraph is close to its expectation we use Talagrand's concentration inequality [9]. In its usual form Talagrand's inequality is asymmetric and bounds a random variable in terms of its median. We use the following symmetric version (see [7, Chapter 20]) that gives concentration of the random variable about its mean.

Theorem 9. *Let $\Omega = \prod_{i=1}^{N} \Omega_i$ be a product of probability spaces with the product measure. Let X be a random variable on Ω such that*

(i) $|X(\omega) - X(\omega')| \leq c$ whenever ω and ω' differ on only a single coordinate for some constant $c > 0$;

(ii) whenever $X(\omega) \geq r$ there is a set $I \subseteq \{1, \dots, N\}$ with $|I| = r$ such that $X(\omega') \geq r$ for all $\omega' \in \Omega$ with $\omega'_i = \omega_i$ for all $i \in I$.

Then for $0 \leq s \leq \mathbb{E}(X)$,

$$\mathbb{P}\left(|X - \mathbb{E}(X)| \geq s + 60c\sqrt{\mathbb{E}(X)}\right) \leq 4e^{-s^2/8c^2\mathbb{E}(X)}.$$

Lemma 10. *Let G be an n-vertex ϵ-quasirandom graph with $\rho\binom{n}{2}$ edges and let W be a random walk on G of length βn^2. Then with probability at least $1 - o(1)$, for all $A, B \subseteq V(G)$ with $e_G(A,B) \gg n$, we have $e_{G_{list}(\nu)}(A,B) = (e^{-2\nu} + o(1))e_G(A,B)$.*

Proof. We apply Theorem 9 to the space $\Omega = \prod_{v \in V(G)} \prod_{i=1}^{\nu d(v)} N(v)$, where each neighbourhood has the uniform probability measure; we can view Ω as the space of choices for the first $\nu d(v)$ entries of each list L_v. Let $A, B \subseteq V(G)$ be sets in G with $e_G(A,B) \gg n$. Let G_{rem} denote the subgraph of G consisting of those edges which are removed from G to obtain $G_{list}(\nu)$ and consider the random variable $X_{A,B} = e_{G_{rem}}(A,B)$. It is easy to see that $X_{A,B}$ satisfies the conditions of Talagrand's inequality. Indeed, (i) holds since changing a list entry can change $X_{A,B}$ by at most $c = 2$. Furthermore, (ii) holds since if $X_{A,B} \geq s$, there are s list entries witnessing this fact. Therefore, by Theorem 9, for $120\sqrt{\mathbb{E}(X_{A,B})} \leq t \leq \mathbb{E}(X_{A,B})$ we have

$$\mathbb{P}(|X_{A,B} - \mathbb{E}(X_{A,B})| \geq 2t) \leq 4e^{-t^2/32\mathbb{E}(X_{A,B})}.$$

But by Lemma 8 we have $\mathbb{E}(X_{A,B}) = (1 - e^{-2\nu} + o(1))e_G(A,B) \gg n$. Taking $t = C'\sqrt{n\mathbb{E}(X_{A,B})}$ $(= o(\mathbb{E}(X_{A,B})))$ for large enough $C' > 0$ gives that

$$\mathbb{P}\left(\left|X_{A,B} - (1 - e^{-2\nu} + o(1))e_G(A,B)\right| \geq 2t\right) \leq 8^{-n}.$$

But there are at most 2^n choices for A and 2^n choices for B. Therefore, with probability at least $1 - 2^{-n}$ we have $X_{A,B} = (1 - e^{-2\nu} + o(1))e_G(A,B)$, for all pairs (A,B) with $A, B \subseteq V(G)$ and $e_G(A,B) \gg n$. Since $e_{G_{list}(\nu)}(A,B) = e_G(A,B) - X_{A,B}$ the result follows. □

3 General case

We now move to the case of a general ϵ-quasirandom graph G with edge density ρ. Such G must always contain a connected component of order at least $(1-\epsilon)n$ (as otherwise we can find two

large sets with no edges between them), so by restricting our walk to this component we can assume that G is connected.

The extra difficulty in the general case is that there might be small sets of vertices that are only weakly connected to the rest of the graph in which the random walk can get 'stuck'. For example, let G be a graph consisting of a small clique of order $\epsilon^2 n/2$ joined to a large clique of order $(1-\epsilon^2/2)n$ by a single edge. Then G is ϵ-quasirandom, but it is not even true that the number of edges in G' is concentrated near some value. Indeed, if we start our random walk in the large clique then with positive probability (depending on ϵ but not on n) W will lie entirely within the large clique, but there is also positive probability (depending on ϵ but not on n) that W will cross to the small clique in the first ϵn^2 steps and remain there. So for general quasirandom graphs we cannot hope for as strong a result as Theorem 1, and our assertions about high probability will necessarily depend on ϵ as well as n. In this section we use 'with high probability' to mean 'with probability $1-o_\epsilon(1)$', with $o_\epsilon(1)$ small (depending on ϵ) for large n as defined in Section 1.

Our task in this section is to find a weaker replacement for Corollary 7 in Section 2.1. Recall that we call a vertex v *balanced* if $|d(v)-\rho n| \le \epsilon n$. We will show that if W is a random walk of length βn^2 on G with W_0 balanced, then with high probability W hits most vertices of G about the right number of times. The results in Section 2.2 can then be used to prove Theorem 2 in the same way that Theorem 1 was deduced from Corollary 7.

Our first lemma gives a lower bound on the probability that a given step of a random walk W is in a set $S \subset V(G)$. Write $\mathbf{1}_X$ for the indicator function of a set X and $\mathbf{1}_v$ for the indicator function of the set $\{v\}$. Note that if the initial distribution for W_0 is π then $\mathbb{P}(W_i \in S) = \sum_{v \in S} \pi_v = \pi \cdot \mathbf{1}_S$ for any set $S \subset V(G)$ when $i \ge 0$. The next result shows that this is still almost true if W starts from a balanced vertex, S is large and $i \ge 2$.

Lemma 11. *Let G be a connected n-vertex ϵ-quasirandom graph with $\rho\binom{n}{2}$ edges and v be a balanced vertex. Let $S \subseteq V(G)$ with $|S| \ge \epsilon n$. Then for a random walk W starting at v we have*

$$\mathbb{P}(W_i \in S) \ge \pi \cdot \mathbf{1}_S - 8\sqrt{\epsilon}/\rho \ge |S|/n - 9\sqrt{\epsilon}/\rho,$$

for $i \ge 2$ and n sufficiently large.

Proof. We first show that the random walk is quite well mixed after only two steps. Let A be the set of neighbours of v with degree at most $(\rho+\epsilon)n$ and B be the set of vertices with at least $(\rho-\epsilon)|A|$ neighbours in A — the 'well-behaved' first and second neighbourhoods of v. By ϵ-quasirandomness, $|A| \ge d(v) - \epsilon n \ge (\rho - 2\epsilon)n$ and $|B| \ge (1-\epsilon)n$. We have

$$\mathbf{1}_v P = \frac{1}{d(v)} \mathbf{1}_{N(v)} \ge \frac{1}{(\rho+\epsilon)n} \mathbf{1}_A,$$

where the inequality holds in each coordinate. For $x \in B$,

$$(\mathbf{1}_A P)_x = \sum_{\substack{y \in A \\ xy \in E(G)}} \frac{1}{d(y)} \ge \frac{(\rho-\epsilon)(\rho-2\epsilon)n}{(\rho+\epsilon)n} \ge \rho(1-4\epsilon/\rho),$$

where the first inequality holds since each $y \in A$ has degree at most $(\rho+\epsilon)n$, x has $(\rho-\epsilon)|A|$ neighbours in A and $|A| \geq (\rho - 2\epsilon)n$. Since the entries of P are non-negative we can compose these inequalities to obtain

$$\mathbf{1}_v P^2 \geq \frac{(1-5\epsilon/\rho)}{n}\mathbf{1}_B.$$

Let $\mathbf{b} = \frac{(1-5\epsilon/\rho)}{n}\mathbf{1}_B$. Since $\pi_x = \frac{d(x)}{2\rho\binom{n}{2}}$, if x is a balanced vertex then $\frac{(1-\epsilon/\rho)}{n-1} \leq \pi_x \leq \frac{(1+\epsilon/\rho)}{n-1}$; otherwise we have the weaker bound $\pi_x \leq \frac{1}{\rho n}$. Since at most $2\epsilon n$ vertices are unbalanced and at most ϵn vertices are not in B,

$$\|\mathbf{b}-\pi\|_2 \leq \left(n\left(\frac{7\epsilon}{\rho n}\right)^2 + 3\epsilon n\left(\frac{2}{\rho n}\right)^2\right)^{1/2} \leq \left(\frac{64\epsilon}{\rho^2 n}\right)^{1/2}.$$

Then, for $i \geq 2$,

$$\begin{aligned}
\mathbb{P}\left(W_i \in S\right) &= \mathbf{1}_v P^i \mathbf{1}_S \\
&= \mathbf{1}_v P^2 \cdot P^{i-2}\mathbf{1}_S \\
&\geq \mathbf{b}P^{i-2}\mathbf{1}_S \\
&= \pi P^{i-2}\mathbf{1}_S + (\mathbf{b}-\pi)P^{i-2}\mathbf{1}_S.
\end{aligned}$$

By Cauchy-Schwarz, and the fact that the eigenvalues of P are at most 1,

$$\|(\mathbf{b}-\pi)P^{i-2}\mathbf{1}_S\|_2 \leq \|\mathbf{b}-\pi\|_2\,\|\mathbf{1}_S\|_2 \leq \left(\frac{64\epsilon|S|}{\rho^2 n}\right)^{1/2} \leq 8\sqrt{\epsilon}/\rho,$$

and so

$$\mathbb{P}\left(W_i \in S\right) \geq \pi \cdot \mathbf{1}_S - 8\sqrt{\epsilon}/\rho,$$

proving the first inequality. Since at least $|S| - 2\epsilon n$ elements of S are balanced,

$$\pi \cdot \mathbf{1}_S = \sum_{x \in S}\frac{d(u)}{2\rho\binom{n}{2}} \geq \frac{(|S|-2\epsilon n)(\rho-\epsilon)}{\rho n} \geq |S|/n - 2\epsilon - \epsilon/\rho \geq |S|/n - \sqrt{\epsilon}/\rho,$$

which proves the second inequality. $\square$

We now consider the following variant of the list model for constructing a random walk. Fix some small length L and let $K = \beta n^2/L$. By a *block rooted at* v we mean a random walk of length L starting at v. For each vertex v, let Λ_v be an infinite list of blocks rooted at v. We construct a random walk of length βn^2 as follows. Choose W_0 from the given initial distribution, and, at each stage $s = 1, \ldots, K$, let $W_{(s-1)L} \cdots W_{sL}$ be the first unused block rooted at $W_{(s-1)L}$. At the end of the construction we have examined K blocks in total from the top of the n lists. Let M be the set of blocks examined (equivalently, the *multiset* of roots of blocks used).

This construction generalises the simple list model (which corresponds to the case $L = 1$), and we again hope to exploit the independence of blocks by applying standard concentration inequalities. There are two main obstacles. One is that we do not know anything about the distribution of a block rooted at a vertex v which is not balanced. We therefore first show that most of the root

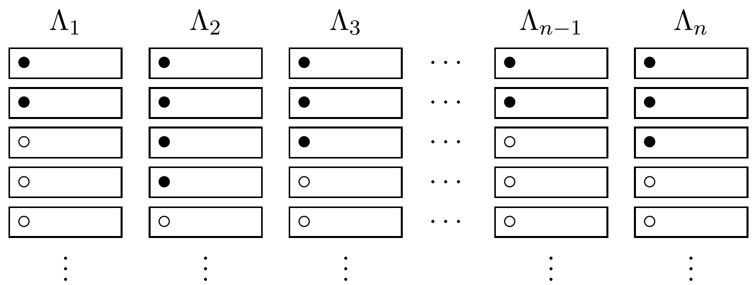

Figure 1: The construction examines K blocks from the top of the lists Λ_v, but we cannot tell in advance which blocks these will be.

vertices are balanced. The second obstacle is that we do not know in advance which set of blocks we will examine. We handle this by approaching the problem from the other direction: for a given multiset M, what is the probability that the corresponding blocks do not contain an even distribution of the vertices? This turns out to be small enough that summing over all possible M gives the bound we require.

Lemma 12. *Let G be a connected n-vertex ϵ-quasirandom graph with $\rho\binom{n}{2}$ edges and let W be a random walk of length βn^2 starting at a balanced vertex of G. Let $\delta = 3\sqrt[4]{\epsilon}/\sqrt{\rho}$ and suppose that n is sufficiently large. Then with probability at least $1-3\delta$ there exists a set $B \subseteq V(G)$ with $|B| \geq (1-\delta)n$ such that each vertex in B is hit at least $(1-4\delta)\beta n$ times by W.*

Proof. Take $L = \omega(n)$ for any $\omega(n) \ll n/\log n$ which tends to infinity as $n \to \infty$, and let $K = \beta n^2/L$. Construct a random walk W as described above and let $x_1, \dots, x_K$ be the roots of the K blocks used. We first show that with high probability many of the vertices $\{x_1, \dots, x_K\}$ are balanced.

Let U be the number of x_i that are unbalanced. By Lemma 11, for $i \geq 2$,

$$\mathbb{P}\left(x_i \text{ is unbalanced}\right) \leq 1 - \left((1-2\epsilon) - \delta^2\right) \leq 2\delta^2,$$

since there are at least $(1-2\epsilon)n$ balanced vertices and $\delta^2 > 2\epsilon$. By Markov's inequality,

$$\mathbb{P}\left(U \geq \delta K\right) \leq \frac{\mathbb{E}\left(U\right)}{\delta K} \leq \frac{2\delta^2 K}{\delta K} = 2\delta.$$

Now let M be a multiset of $(1-\delta)K$ balanced vertices and let $W^{(1)}, W^{(2)}, \dots, W^{((1-\delta)K)}$ be the corresponding blocks. We will show that the probability that these blocks contain most balanced vertices about the right number of times is large.

Let $S \subseteq V(G)$ with $|S| \geq \delta n$. By Lemma 11, for every $1 \leq i \leq (1-\delta)K$ and every $j \geq 2$ we have $\mathbb{P}\left(W_j^{(i)} \in S\right) \geq \delta - \delta^2$. Let X_{ij} be the indicator of the event $W_j^{(i)} \in S$, let $X_j = \sum_{i=1}^{K} X_{ij}$ and let $X_{M,S} = \sum_{j=1}^{L} X_j$. For fixed j the X_{ij} are independent, so by Chernoff's inequality (see [1, Appendix A]),

$$\mathbb{P}\left(X_j < (\delta - 2\delta^2)|M|\right) \leq e^{-2\delta^4|M|}.$$

Hence

$$\begin{aligned}\mathbb{P}\left(X_{M,S} < (\delta - 4\delta^2)\beta n^2\right) &\leq \mathbb{P}\left(X_{M,S} < (\delta - 3\delta^2)(1-\delta)KL\right) \\ &\leq \mathbb{P}\left(X_j < (\delta - 2\delta^2)|M| \text{ for some } 2 \leq j \leq L\right) \\ &\leq Le^{-2\delta^4|M|},\end{aligned}$$

where the second inequality holds for large n because the contribution from X_1 is negligible as $L \to \infty$.

If the random walk W fails to hit at least $(1-\delta)n$ vertices at least $(1-4\delta)\beta n$ times each then either δK of the x_i are unbalanced or there is an M and an S such that $X_{M,S} < (\delta - 4\delta^2)\beta n^2$. But the probability of this bad event is at most

$$\begin{aligned}\mathbb{P}(U \geq \delta K) + \sum_M \sum_S Le^{-2\delta^4|M|} &\leq 2\delta + \binom{K+n-1}{n-1}\binom{n}{\geq \delta n} Le^{-2\delta^4(1-\delta)K} \\ &\leq 2\delta + O(K)^n \cdot 2^n \cdot L \cdot e^{-2\delta^4(1-\delta)K} \\ &\leq 2\delta + \exp\bigl(O(n\log n) + O(n) + O(\log n) - 2\delta^4(1-\delta)K\bigr) \\ &\leq 3\delta,\end{aligned}$$

for n sufficiently large, since $K \gg n \log n$. □

We now have everything we need to complete the proof of Theorem 2.

Proof of Theorem 2. We will show that with probability $1 - o_\epsilon(1)$ the graph G' obtained from G by removing the edges of W is close to $G_{list}(\beta/\rho)$. It then follows from Lemma 10 that G' is η-quasirandom with probability $1 - o_\epsilon(1)$.

Since there are at most $2\epsilon n < \delta$ unbalanced vertices in G, by Lemma 12 with probability at least $1 - 3\delta$ there is a set B of $(1-2\delta)n$ balanced vertices such that every $v \in B$ is hit at least $(1-4\delta)\beta n \geq (1-5\delta)\frac{\beta}{\rho}d(v)$ times by W. This accounts for $(1-2\delta)n \cdot (1-4\delta)\beta n \geq (1-7\delta)\beta n^2$ of the list entries examined, so G' differs from $G_{list}(\beta/\rho)$ by at most $14\delta\beta n^2$ edges. Since δ tends to 0 with ϵ, the result follows. □

4 Trees

A *homomorphism* from a graph H to a graph G is an edge-preserving map $\phi : V(H) \to V(G)$. A random walk can be viewed as a random homomorphism of a path; a natural generalisation is to consider a random homomorphism of some other tree T.[1] Just as we traversed a path in one direction, our trees will be rooted and we think of them as directed 'downwards', away from the root. In this section we will explore to what extent the methods of Section 3 can be applied in this more general setting.

We generate a random homomorphism as follows. Enumerate the vertices of T as $v_0, v_1, \ldots, v_k$ where, for each j, $T[v_0, \ldots, v_j]$ is a connected subtree of T containing the root v_0. First choose

[1] Sometimes called a *tree-indexed random walk*.

$\phi(v_0)$ from a given initial distribution. Then, at each stage $j > 0$, let u be the parent of v_j in T and choose $\phi(v_j)$ uniformly at random from the neighbours of $\phi(u)$. All choices are made independently, and we can think of these choices as being taken from the lists L_v as before.

Suppose now that G is an ϵ-quasirandom graph on n vertices. Let ϕ be a random homomorphism of a tree T of size βn^2 to G, and let G' be the graph obtained from G by deleting the edges of $\phi(T)$. Is G' quasirandom with high probability? It is easy to see that in general the answer is no. For example, let $G = K_n$ and T be an $n/2$-ary tree of depth 2 (here $\beta = 1/4 + o(1)$). Then with high probability $\phi(T)$ contains a constant fraction of the edges of G. But all of these edges are incident on the neighbourhood of the root, which has $(1 - e^{-1/2} + o(1))n$ vertices with high probability, so with high probability G' is not quasirandom.

We seek conditions on T such that we can apply the approach taken in Section 3 with minimal changes. The condition we give here imposes an upper bound on the maximum degree of T.

We need an analogue of the second model for the construction of a random walk. Instead of breaking our path into many short paths, we break our tree into many small edge-disjoint subtrees.

Lemma 13. *Let T be a rooted tree with N edges and let $L \leq N$. Then T can be written as an edge-disjoint union of rooted trees $R_1, \ldots, R_K$, each of size between L and $3L$.*

Proof. Let v be a vertex of T furthest from the root such that v has at least L descendants. Then each branch of T lying below v has at most L edges, so some union of these branches has size between L and $2L$; let this be R_1. We obtain $R_2, \ldots, R_K$ similarly until there are less than L edges of T remaining, which we add to R_K. □

Write $\mathcal{R} = \{R_1, \ldots, R_K\}$ for the corresponding set of abstract rooted trees, up to isomorphism. In an abuse of notation we use R_i to refer to both the specific subtree of T and its isomorphism type.

It is convenient to number the R_i such that $R_1 \cup \cdots \cup R_j$ is a subtree of T containing the root for each j. We can then describe the second model for the construction of a random homomorphism as follows. For each $v \in V(G)$ and $R \in \mathcal{R}$, let $\Lambda_{v,R}$ be a list of independent random homomorphisms from R to G that map the root of R to v. Choose a vertex v_1 from the given distribution for the image of the root of T and identify $\phi(R_1)$ with the first entry from Λ_{v_1,R_1}. (If R_1 has a non-trivial automorphism group then there is a choice of identification of R_1 with the reference copy in $\mathcal{R}$. The choice is unimportant provided the same choice is made every time.) Then at each stage j we have already determined the image v_j of the root of R_j, and we identify $\phi(R_j)$ with the first unused element from Λ_{v_j,R_j}.

Now let T be a rooted tree with βn^2 edges. As before we want to show that T 'visits' most vertices of G about the right number of times. We need to be careful here about what counts as a 'visit': what we want to count is the number of times an edge leaves a vertex, as that is the number of entries of the corresponding list that will be examined. So we say $\phi(T)$ *visits* $x \in V(G)$ whenever uv is an edge of T with u the parent of v and $\phi(u) = x$; the number of visits $\phi(T)$ makes to x is the number of edges uv for which this occurs.

There are three places where the argument in the proof of Lemma 12 needs modification or additional details need to be checked.

(i) In the path case the edges (or vertices) of the blocks had a natural order and the blocks were all the same size. In the tree case we are free to choose a labelling of the edges in each block, but the blocks might still have different sizes: when we look at the $2L$th edge from each block, are there enough blocks with $2L$ edges that Chernoff's inequality will give good concentration?

(ii) In the path case the set of list entries examined was parameterised by multisets of vertices of G. In the tree case the set of list entries examined is instead parameterised by multisets of pairs (v, R) with $v \in V(G)$ and $R \in \mathcal{R}$. So the factor $\binom{K+n-1}{n-1}$ in the final sum needs to be replaced by $\binom{K+n|\mathcal{R}|-1}{n|\mathcal{R}|-1}$, and we must restrict the size of $|\mathcal{R}|$ to prevent this becoming too large.

(iii) In the path case we had to ignore the first two vertices of each block as we needed to take two steps before we had good information about the distribution over vertices. This was safe because the ignored vertices were only a $o(1)$ fraction of the total number of vertices. In the tree case we must ignore the edges whose start point is the root of the block or is a child of the root. We need to ensure that the number of ignored edges is at most a small fraction of the total number of edges.

Problem (i) is avoided by throwing away the small number of edges that receive a label shared by few other edges. If we throw away all edges that receive a label which is used less that $\epsilon n^2/L^2$ times then the total number of edges thrown away is less than $3\epsilon n^2/L$ as there are at most $3L$ edges in each block.

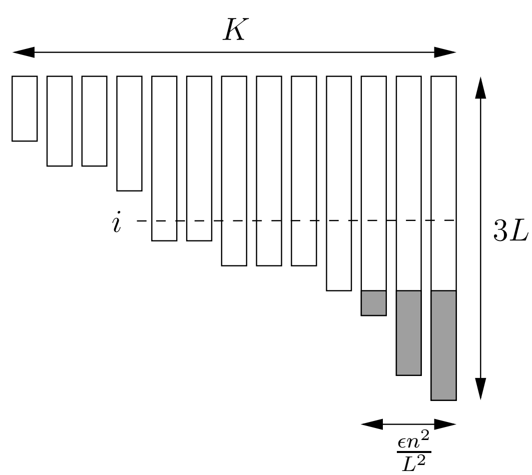

Figure 2: Deleting a $o(1)$ fraction of the edges ensures that the remaining labels i are each used in a large number of blocks.

Problem (ii) is avoided by taking L small: $L = \frac{\log n}{2\log 3}$ suffices. Indeed, since the number of rooted trees on L vertices is $O((2.9955\ldots)^L)$ (see [8]) and $\frac{\beta n^2}{3L} \le K \le \frac{\beta n^2}{L}$, we have in this case that $n|\mathcal{R}| \ll n^{3/2} \ll K$, and

$$\binom{K+n|\mathcal{R}|-1}{n|\mathcal{R}|-1} \ll K^{n|\mathcal{R}|} \ll \exp\left(O(n^{3/2}\log n)\right),$$

which is small enough that it will not overpower the e^{-cK}-type decay.

Problem (iii) is avoided by having Δ^2, the square of the maximum degree of T small (depending on the desired level of quasirandomness) compared to L: so Δ can be as large as a small multiple of $\sqrt{\log n}$.

With these modifications to our earlier argument we obtain the following result.

Theorem 14. *Given $\beta, \rho, \eta > 0$ there exists $\epsilon, c > 0$ such that the following holds. Let G be an n-vertex ϵ-quasirandom graph with $\rho\binom{n}{2}$ edges, T be a rooted tree of size βn^2 with maximum degree $\Delta \leq c\sqrt{\log n}$ and let ϕ be a random homomorphism from T to G such that the image of the root is balanced. Then, with probability $1 - o_\epsilon(1)$, the graph G' formed by removing the edges of $\phi(T)$ from G is η-quasirandom with $(e^{-2\beta/\rho} + o_\epsilon(1))\rho\binom{n}{2}$ edges.*

It would be interesting to know how large $\Delta(T)$ can be taken in Theorem 14. By the example at the start of this section we must have $\Delta(T)$ small compared to n. Is this already enough?

Acknowledgements. We would like to thank the organisers of the *Probabilistic methods in graph theory* workshop at the University of Birmingham where we heard about this problem. We would also like to thank Jan Hladký for some helpful discussions and for suggesting the extension of our results on paths to more general trees.

References

[1] N. Alon and J. Spencer. *The probabilistic method.* John Wiley & Sons Inc., third edition, 2008.

[2] F. R. K. Chung, R. L. Graham, and R. M. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, 1989.

[3] J. Hladký. Problem session, Probabilistic methods in graph theory, University of Birmingham, 2012.

[4] M. Jerrum and A. Sinclair. Approximate counting, uniform generation and rapidly mixing Markov chains. *Inform. and Comput.*, 82(1):93–133, 1989.

[5] M. Krivelevich and B. Sudakov. Pseudo-random graphs. In *More sets, graphs and numbers*, volume 15 of *Bolyai Soc. Math. Stud.*, pages 199–262. Springer, Berlin, 2006.

[6] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times.* American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.

[7] M. Molloy and B. Reed. *Graph colouring and the probabilistic method.* Springer-Verlag, 2002.

[8] Richard Otter. The number of trees. *Ann. of Math. (2)*, 49:583–599, 1948.

[9] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.*, (81):73–205, 1995.

[10] Andrew Thomason. Pseudorandom graphs. In *Random graphs '85 (Poznań, 1985)*, volume 144 of *North-Holland Math. Stud.*, pages 307–331. North-Holland, Amsterdam, 1987.